NSWC TR 82-49

WEPOR:
A WEIGHTED POLYNOMIAL REGRESSION PROGRAM

BY PATRICIA A. SHIELDS MARLIN A. THOMAS

STRATEGIC SYSTEMS DEPARTMENT

SEPTEMBER 1982

Approved for public release; distribution unlimited.

NAVAL SURFACE WEAPONS CENTER

Dahlgren, Virginia 22448 • Silver Spring, Maryland 20910

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NSWC TR 82-49	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) WEPOR: A WEIGHTED POLYNOMIAL REGRESSION PROGRAM		5. TYPE OF REPORT & PERIOD COVERED Final 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Patricia A. Shields Marlin A. Thomas		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Surface Weapons Center (K106) Dahlgren, VA 22448		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Surface Weapons Center Dahlgren, VA 22448		12. REPORT DATE September 1982 13. NUMBER OF PAGES 35
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Weighted Regression
Polynomial Regression
Regression
Weighted Least Squares
Lease Squares Estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two weighted least squares regression programs are documented and exemplified. Both programs apply to polynomial models with one independent variable. Program WEPOR was formulated for the case in which the error terms have different variances but are uncorrelated, whereas WEPOR2 deals with the problem of different variances and correlated error terms. Output from both programs include ANOVA (ANalysis Of VAriance) tables, predicted values of the dependent variable and the associated residuals, confidence and prediction limits for selected synthetic points, and a plot of the sample points, the regression curve, and the confidence and prediction limits.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

FOREWORD

The work described herein was performed by the Mathematical Statistics Staff (K106), Space and Surface Systems Division, Strategic Systems Department. It was motivated by a request from Mr. Donald R. Monn of the Weapons Systems Department to design and analyze a projectile seating distance (PSD) experiment. The authors wish to acknowledge several stimulating discussions with Mr. Monn regarding the application of weighted polynomial regression to the PSD experiment. The date of completion was September 1981.

This report was reviewed by Mr. Carlton W. Duke, Head, Space and Surface Systems Division.

Released by:

Oliver F. Braxton

OLIVER F. BRAXTON, Head
Strategic Systems Department

CONTENTS

Page

INTRODUCTION

Polynomial regression is a methodology used to fit curvilinear models to a set of observations. These curvilinear models fit into the framework of the general linear model and, hence, can usually be fit to the data using any general multiple regression program. Two such programs are currently available through the Mathematical Statistics Staff, viz. GEMREG (GEneral Multiple REGession)* and DA-MRCA (DAhlgren Multiple Regression and Correlation Analysis).** Both provide least squares estimates of the regression coefficients, analysis of variance tables, and a variety of user-controlled options. These programs hinge on the assumptions that the error terms (differences between the observed and predicted values of the dependent variable) can be assumed to have zero expectation, the same variance for all observations, and zero correlation. When these last two assumptions for the error terms are not met, the usual least squares method is not applicable; instead, a weighted least squares procedure is required.

Programs WEPOR and WEPOR2 (WEighted POlynomial Regression) use this weighted least squares procedure to estimate regression coefficients for models with one independent variable. Program WEPOR handles the case in which the error terms have different variances but are uncorrelated, whereas WEPOR2 deals with the problem of different variances and correlated error terms. Output for both programs includes ANOVA (ANalysis Of VAriance) tables, predicted values of the dependent variable and the associated residuals, and confidence limits for selected synthetic points. The values for bounds on the entire curve generated from the input data are written on output files for use with DISSPLA (Display Integrated Software System and Plotting LAnguage).† An example of a program that uses the output from WEPOR and DISSPLA features to plot sample points, the regression curve, and confidence and prediction limits is program LIMITS.

---

*Taub, A. E., and M. A. Thomas, *GEMREG - A General Multiple Regression Program*, NSWC TN 81-298, (Dahlgren, Va., 1981).

**Abt, K., G. Gemmill, T. Herring, and R. Shade, *DA-MRCA: A Fortran IV Program for Multiple Linear Regression*, NSWC TR-2035, (Dahlgren, Va., 1966).

†Integrated Software Systems Corporation (ISSCO), *Display Integrated Software System and Plotting Language*, ISSCO (San Diego, Calif., 1970).

THE MODEL

The polynomial regression model with a single variable has form

$$y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_k X_i^k + e_i, \qquad i = 1,2, \ldots n. \tag{1}$$

In this model, X_i is the value of the independent variable associated with the ith response value (y_i), n is the number of observations, k is the order of the polynomial, β_j is the jth regression coefficient, and e_i is the ith random error. The inclusion of e in the model accounts for the fact that the response variable y is a random variable and, hence, the relationship between the response variable and the independent variable is not an exact functional relationship.

Polynomial models fit into the framework of the general linear model

$$y_i = \beta_o + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + e_i, \qquad i = 1,2, \ldots n \tag{2}$$

and, hence, can usually be fit to data using any general multiple regression program provided that the e_i can be assumed to have zero expectation, the same variance σ^2 for all i, and be uncorrelated. These assumptions can be expressed in a more compact form if the model is written in matrix notation:

$$\underline{y} = X\underline{\beta} + \underline{e}. \tag{3}$$

In the general context of model (2), $\underline{y}$ is an n x 1 vector of observations, $\underline{\beta}$ is a (k + 1) x 1 vector of regression coefficients, $\underline{e}$ is an n x 1 vector of random errors, and

$$\underset{n \times (k+1)}{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} .$$

Since model (1) is a special case of model (2), the appropriate X for (1) is obtained by letting $X_{ki} = X_i^k$. In this notation, the expectation of the e_i and their variance-covariance matrix can be denoted by $E(\underline{e})$ and $Var(\underline{e})$, respectively. Hence, if the e_i are assumed to have zero expectation, this is denoted by $E(\underline{e}) = \underline{0}$. Also if the e_i are assumed to be uncorrelated with the same variance, this is denoted by $Var(\underline{e}) = \sigma^2 I$ where I is the n x n identity matrix. In regression applications, the assumption $E(\underline{e}) = \underline{0}$ does not present any difficulty. However, the assumption $Var(\underline{e}) = \sigma^2 I$ cannot always be met and, hence, poses a serious problem if not handled properly. In this case, the variance-covariance matrix is denoted by $Var(\underline{e}) = \sigma^2 V$ where V is an n x n positive definite matrix.

An example of a regression application where $Var(\underline{e}) \neq \sigma^2 I$ involves regressing projectile seating distances (y) on given barrel life (x) expressed in percent expended. Here, the variation in seating distance increases with the percent expended barrel life. Hence, the assumption of equal variances does not hold, and the usual least squares regression is not applicable. Cases of this kind and more complicated situations where the errors are correlated can be handled by a modified least squares procedure known as *weighted* least squares. This procedure is discussed by Draper and Smith,* and much of the development that follows is based on their discussion.

When the aforementioned assumptions are satisified, the usual least squares procedure provides a vector of estimates of the regression coefficients that has the form

$$\underline{\hat{\beta}} = \underline{b} = (X'X)^{-1} X'\underline{y}. \tag{4}$$

The weighted least squares procedure amounts to transforming the dependent or response variable y to another variable that does satisfy the assumptions. The usual (unweighted) least squares analysis is then applied to the new variable, and the estimates so obtained are reexpressed in terms of the original variable y. This process is examined in details in the ensuing paragraphs.

Consider the original model (Equation 3) with assumptions $E(\underline{e}) = \underline{0}$ and $Var(\underline{e}) = \sigma^2 V$ (vice $\sigma^2 I$). Since V is positive definite, it is possible to find an upper triangular matrix P such that $P'P = V$. (Draper and Smith indicate that it is possible to find a unique nonsingular symmetric matrix P such that $P'P = PP = P^2 = V$. We have not found this to be the case, nor is such a requirement necessary in what follows.)

---

*Draper, N. R. and H. Smith, *Applied Regression Analysis* (New York, N.Y.: John Wiley & Sons, Inc., 1966), pp. 77-81.

If the model in Equation 3 is premultiplied by $(P')^{-1}$, a new model is generated in the form

$$(P')^{-1} \underline{y} = (P')^{-1} X \underline{\beta} + (P')^{-1} \underline{e} \tag{5}$$

Since

$$E\,[(P')^{-1}\underline{e}] = (P')^{-1} E(\underline{e}) = (P')^{-1}\, \underline{0} = \underline{0}$$

and

$$\begin{aligned} \mathrm{Var}\,[(P')^{-1}\underline{e}] = E[(P')^{-1}\, \underline{e}\underline{e}'\, P^{-1}] &= (P')^{-1}\, E(\underline{e}\, \underline{e}')P^{-1} \\ &= (P')^{-1}\, V\, P^{-1}\, \sigma^2 \\ &= (P')^{-1}\, (P'P)P^{-1}\, \sigma^2 \\ &= I\sigma^2, \end{aligned}$$

the new model meets the assumptions required for the ordinary least squares procedure. This new model can be written in matrix notation as

$$\underline{z} = Q\, \underline{\beta} + \underline{f} \tag{6}$$

where $\underline{z} = (P')^{-1}\, \underline{y}$, $Q = (P')^{-1}\, X$, and $\underline{f} = (P')^{-1}\, \underline{e}$.

THE ANALYSIS

The error term $\underline{f}$ in the revised model in Equation 6 satisfies the assumptions for the usual least squares analysis. Therefore, the usual analysis will be applied to the revised model. Estimation of the regression coefficients will be dealt with first. These estimates are obtained by writing the solution vector in Equation 4 in terms of the new parameters in Equation 6. This provides

$$\underline{b} = (Q'Q)^{-1}\ Q'\ \underline{z} \tag{7}$$

Reexpressing Q and $\underline{z}$ in terms of the original model parameters provides

$$\begin{aligned} \underline{b} &= [(X'P^{-1})((P')^{-1}\ X)]^{-1}\ (X'P^{-1})((P')^{-1}\ \underline{y}) \\ &= [X'\ (P'P)^{-1}\ X]^{-1}\ X(P'P)^{-1}\ \underline{y} \\ &= (X'\ WX)^{-1}\ X'\ W\underline{y}\ . \end{aligned} \tag{8}$$

In this expression, W is the inverse of V; i.e., $W = (P'P)^{-1} = V^{-1}$.

This solution has the same form as Equation 4, except for the insertion of W, the weighting matrix. The new model has an implied zero intercept, since the Q matrix does not have a leading column of ones. Hence, the entries in the analysis of variance table for the new model are computed in a slightly different manner from those obtained when ordinary least squares procedures are used. Table 1 shows the breakdown of the degrees of freedom and formulae needed to compute the sums of squares for a first degree polynomial.

Table 1. Analysis of Variance Table for First Degree Polynomial

Source	Sums of Squares	Degrees of Freedom
β_o	$(\Sigma(q_o)_i\, z_i)^2/\Sigma(q_o)_i^2$	1
$\beta_1 \mid \beta_o$	$\underline{b}'X'W\underline{y} - SS(\beta_o)$	1
Error	$\underline{y}'\, W\underline{y} - \underline{b}'X'W\underline{y}$	n - 2
Total	$y'Wy\ (= \underline{z}'\, \underline{z})$	n

In this table, $(q_o)_i$ is the ith element of the first column in the Q matrix $[Q = (P')^{-1}\ X]$.

When several observations are taken at the same level of the independent variable, the error sum of squares in Table 1 can be broken into components for lack of fit and pure error. From Equations 5 and 6, we have

$$\underline{z} = (P')^{-1}\ \underline{y}$$

where $z' = (z_1, z_2, \ldots z_n)$. A change in the subscript of the z's produces

$$z' = (z_{11}, z_{12}, \ldots z_{1n_1}, z_{21}, \ldots z_{2n_2}, \ldots z_{ki}, z_{k2}, \ldots z_{kn_k})$$

where the first n_1 values are associated with the first level of the independent variable, the next n_2 values are associated with the second level, and so on. With this notation, the sum of squares for pure error is computed by

$$SS\ (pe) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2$$

$$= \sum_{i=1}^{k} \left[\sum_{j=1}^{n_i} (z_{ij})^2 - \frac{\left(\sum_{j=1}^{n_i} z_{ij}\right)^2}{n} \right] \quad (9)$$

with degrees of freedom $\upsilon = \sum_{i=1}^{k} n_i - k$.

The sum of squares for lack of fit can then be obtained by subtraction; i.e.,

$$SS(lf) = SSE - SS(pe)$$

where SSE is the error sums of squares from Table 1.

Confidence limits on the expected value of y and prediction limits on the mean of m future observations of y differ only slightly in weighted regression from unweighted regression. The value of the independent variable X used in forming the limits is referred to as the synthetic point. Letting x^* denote the synthetic point associated with X and $(\underline{x}^*)' = (1, x^*, (x^*)^2, \ldots (x^*)^k)$ for a kth degree polynomial,

$$(\underline{x}^*)' \underline{b} = b_o + b_1 x^* + b_2 (x^*)^2 + \ldots + b_k (x^*)^k$$

is a point estimate of the expected value of y and of a single future observation when $X = x^*$. In the *unweighted* case, the $100(1 - \alpha)$ percent confidence limits on E(y) when $X = x^*$ are

$$(\underline{x}^*)' \underline{b} \pm t_{\upsilon,\ 1-\alpha/2}\ [s^2((\underline{x}^*)' (X'X)^{-1} \underline{x}^*)]^{1/2}\ . \tag{10}$$

In the weighted case, X′X is replaced with X′ WX yielding

$$(\underline{x}^*)' \underline{b} \pm t_{\upsilon,\ 1-\alpha/2}\ [s^2((\underline{x}^*)' (X'WX)^{-1} \underline{x}^*)]^{1/2}\ . \tag{11}$$

In these expressions, $t_{\upsilon,\ 1-\alpha/2}$ is the $100\ (1 - \alpha/2)$ percentage point for a t distribution with υ degree of freedom where υ is associated with the error mean square in the analysis of variance table. This error mean square is denoted by s^2 above and is obtained by dividing the error sums of squares (SSE) by υ, the associated degrees of freedom; i.e., $s^2 = SSE/\upsilon$.

The prediction limits for the mean of m future observations at $X = x^*$ is, in the *unweighted* case, given by

$$(\underline{x}^*)' \underline{b} \pm t_{\upsilon,\ 1-\alpha/2}\ [s^2(\frac{1}{m} + (\underline{x}^*)' (X'X)^{-1} \underline{x}^*)]^{1/2}\ . \tag{12}$$

For the *weighted* case, X′X is changed as above yielding

$$(\underline{x}^*)' \underline{b} \pm t_{\upsilon,\ 1-\alpha/2}\ [s^2(\frac{1}{m} + (\underline{x}^*)' (X'WX)^{-1} \underline{x}^*)]^{1/2}\ . \tag{13}$$

PROGRAM ORGANIZATION

Program WEPOR is actually the main driving routine that calls a series of first level subroutines to perform various tasks.

First, subroutine IOP is called to read in parameters for user-specified input and output options, viz., a title for the execution, the number of observations, the desired degree of the polynomial model to be fit to the data, and a parameter specifying whether or not confidence and prediction limits are requested. The printing of these limits requires the further input of the number of future observations on which the prediction limits are based, and the number of synthetic points to be read in if the levels of x to be used for the limits are different from those in the original data. The validity of each parameter is checked and, should inconsistencies be detected, either a default value is substituted or an error message printed and execution halted.

Subroutine READIN is called to read in the raw input data and the array of weights corresponding to the diagonal elements of the matrix $W = V^{-1}$ (recall that program WEPOR assumes uncorrelated error terms; should correlations exist, program WEPOR2 should be used and the entire matrix V is read in at this point). The required matrix P, where $(P'P)^{-1} = W$, is computed at this time. If V is a diagonal matrix, the elements of P^{-1} are calculated by simply taking the square root of the corresponding elements of W. In the case where V is non-diagonal, P is obtained by performing a matrix decomposition on V using the square root method. The total sum of squares and sum of squares due to pure error are also computed. The raw data points are saved on TAPE10.

For each stage of development in the model, subroutine REGRESS is called to compute and print a set of regression coefficients, the additional sum of squares to be included in the regression sum of squares, and the residual sum of squares and F statistic for the current model.

Subroutine TABLE prints two analysis of variance tables: one shows the breakdown of the residual sum of squares into components of pure error and lack of fit and the other table shows the contribution made by each term in the model to the overall regression sum of squares. This subroutine also prints the raw input data, estimated values for the dependent variable, and residuals.

If the user has requested confidence and prediction limits, subroutine SYNTH performs the necessary calculations. If the user has specified that a new set of levels for x_i are to be used instead of those from the original data set, these new levels are read in from the input file. The confidence and prediction limits are printed and are also saved on TAPE11 and TAPE12, respectively.

Some of these first level subroutines reference routines found in the *NSWC/DL Library of Mathematics Subroutines*:* CROUT, which inverts general real matrices, and MPROD and TMPROD, which perform matrix multiplication operations. Routine QSORT from the *User's Guide for the CDC 6700 Computing System*** is used to arrange the levels of the independent variable in ascending order. Subroutine FINDT, used to estimate the critical t value for confidence and prediction units, is adapted from a similar routine in program GEMREG.†

EXAMPLE

For the application of projectile seating distance (psd) expressed as a function of percent gun barrel life expended, the following independent and uncorrelated pairs of data points were used to derive a first degree polynomial.

Percent Barrel Life Expended	psd (m)
0	39.82
0	39.71
10	41.13
10	41.10
30	43.52
30	43.90
60	48.05
60	46.31
75	47.23
75	48.68

Although the weight associated with each level of the independent variable is the inverse of the variance for the response variable at that level, these variances are unknowns. Estimates based on the above data and data from previous experiments were used to construct the following weights:

*Morrison, Alfred H. Jr., *NSWC/DL Library of Mathematics Subroutines*, NSWC TR 81-410 (Dahlgren, Va., 1981).

***User's Guide for the CDC 6700 Computing System*, NSWC TR-3228 (Dahlgren, Va., 1974).

†Taub, A. E. and M. A. Thomas, 1981.

Value of Independent Variable	Weight
0	12.50
10	10.00
30	5.00
60	1.40
75	1.25

Appendix A provides the input guide for execution of programs WEPOR and WEPOR2. The actual cards used for this example are shown in Appendix B.

Computation of the estimates for the regression parameters requires Equation 8:

$$\underline{b} = (X'WX)^{-1} X'W\underline{y}$$

where

$$X = \begin{bmatrix} 1 & 0. \\ 1 & 0. \\ 1 & 10. \\ 1 & 10. \\ 1 & 30. \\ 1 & 30. \\ 1 & 60. \\ 1 & 60. \\ 1 & 75. \\ 1 & 75. \end{bmatrix}, \quad W = \begin{bmatrix} 12.5 & 0 & . & . & . & . & . & . & . & 0 \\ 0 & 12.5 & . & . & . & . & . & . & . & . \\ . & . & 10.0 & . & . & . & . & . & . & . \\ . & . & . & 10.0 & . & . & . & . & . & . \\ . & . & . & . & 5.0 & . & . & . & . & . \\ . & . & . & . & . & 5.0 & . & . & . & . \\ . & . & . & . & . & . & 1.4 & . & . & . \\ . & . & . & . & . & . & . & 1.4 & . & . \\ . & . & . & . & . & . & . & . & 1.25 & . \\ 0 & . & . & . & . & . & . & . & . & 1.25 \end{bmatrix},$$

and

$$\underline{y} = \begin{bmatrix} 39.82 \\ 39.71 \\ 41.13 \\ 41.10 \\ 43.52 \\ 43.90 \\ 48.05 \\ 46.31 \\ 47.23 \\ 48.68 \end{bmatrix}.$$

The following matrix operations are required before proceeding:

$$(X'WX) = \begin{bmatrix} 60.30 & 855.50 \\ 855.50 & 35142.50 \end{bmatrix}$$

$$(X'WX)^{-1} = \begin{bmatrix} 2.5 \times 10^{-2} & -6.1 \times 10^{-4} \\ -6.2 \times 10^{-4} & 4.35 \times 10^{-5} \end{bmatrix} *$$

$$(X'W\underline{y}) = \begin{bmatrix} 2505.52 \\ 38253.80 \end{bmatrix} .$$

Recall that there exists a matrix P such that $(P'P)^{-1} = W$. The vector $\underline{z} = (P')^{-1}\ \underline{y}$ is then computed

$$\underline{z} = \begin{bmatrix} \sqrt{12.5} & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & \sqrt{12.5} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \sqrt{10.0} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \sqrt{10.0} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \sqrt{5.0} & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{5.0} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{1.4} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{1.4} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{1.25} & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \sqrt{1.25} \end{bmatrix} \begin{bmatrix} 39.82 \\ 39.71 \\ 41.13 \\ 41.10 \\ 43.52 \\ 43.90 \\ 48.05 \\ 46.31 \\ 47.23 \\ 48.68 \end{bmatrix} = \begin{bmatrix} 140.78 \\ 140.40 \\ 130.06 \\ 129.97 \\ 97.32 \\ 98.16 \\ 57.43 \\ 55.35 \\ 52.80 \\ 54.43 \end{bmatrix} .$$

* For the sake cf clarity, rounded values will be given for the results of matrix operations.

Similarly, the vector $\underline{q}_o$, which is equal to the first column in the Q matrix $(= (P')^{-1} X)$, is found to be

$$\underline{q}_o = \begin{bmatrix} \sqrt{12.5} \\ \sqrt{12.5} \\ \sqrt{10.0} \\ \sqrt{10.0} \\ \sqrt{5.0} \\ \sqrt{5.0} \\ \sqrt{1.4} \\ \sqrt{1.4} \\ \sqrt{1.25} \\ \sqrt{1.25} \end{bmatrix}.$$

The vector of estimates for the regression parameters, vector $\underline{b}$, is therefore equal to

$$\underline{b} = (X'WX)^{-1}\ X'W\underline{y} = \begin{bmatrix} 39.88 \\ 0.12 \end{bmatrix}.$$

This result is found on the first page of the printout (Appendix B, page B-4). For the analysis of variance tables printed on pages two and three of the printout (pages B-4, B-5), the following operations are performed:

$$SSR = \underline{b}'\ (X'\ W\underline{y}) = 104424.89$$

$$SS\ (\beta_o) = \left(\sum_{i=1}^{10} (q_o)_i\ z_i\right)^2 / \sum_{i=1}^{10} (q_o)_i^2 = 104106.35$$

$$SS\ (\beta_1 | \beta_o) = SSR - SS(\beta_o) = 318.54$$

$$SS\ Total = \underline{y}'W\underline{y} = \underline{z}'\underline{z} = 104431.63$$

$$SSE = SS\ Total - SSR = 6.74$$

Since each level of the independent variable has two observations associated with it, Equation 9 for computing the sum of squares due to pure error can be written as

$$SS\,(pe) = \sum_{i=1}^{5} \sum_{j=1}^{2} (z_{ij} - z_i)^2$$

and is found to be 3.87. Finally,

$$SS(lf) = SSE - SS(pe) = 2.87 .$$

Page two of the printout (page B-4) shows the analysis of the variance table, including a breakdown of the error sum of squares into the sum of squares due to pure error and lack of fit. The associated degrees of freedom and mean squares are printed for the regression and error terms.

The F statistic to test the lack of fit component [F = MS(1f)/MS(pe)] is 1.23. This result is less than 5.41, the critical F value for $\alpha = 0.05$ with degrees of freedom 3 and 5, and indicates that the first degree polynomial model chosen is not inadequate at the 0.05 level.

On the third page of the printout (page B-5) is an analysis of variance table that shows the contribution made by each term in the model. The sum of squares due to regression is determined for polynominals of degree from 0 (β_o only in model) to the full model chosen. At each stage, the additional sum of squares is computed and stored for use in this table.

The sum of squares represented by X**0 is that associated with the regression model having only β_o in it [$SS(\beta_o)$]. In our example, this value is 104106.35. The nth sum of squares listed, X**n, represents the additional sum of squares obtained by adding β_n to the model that already contains $\beta_o, \beta_1 \ldots \beta_{n-1}$ and can be computed as follows:

$$SS(\beta_n | \beta_o, \beta_1, \beta_2 \ldots \beta_{n-1}) = SS(\beta_o, \beta_1, \ldots \beta_n) - SS(\beta_o, \beta_1, \ldots \beta_{n-1}) .$$

In this example,

$$SS(\beta_1 | \beta_o) = SS(\beta_o, \beta_1) - SS(\beta_o)$$

or 318.54 = 104424.89 - 104106.35.

The column with the heading "F Statistics - MSR/MSE" shows the values of the F test for the model at each stage of development.

The fourth page of the printout lists the case numbers, values x_i of the independent variable, observed values y_i of the dependent variable, the estimated values $\hat{y}_i$ for the dependent variable based on the full regression equation, and the residuals $\hat{y}_i - y_i$ (page B-5).

As in this example, the error terms were presumed to be uncorrelated, which indicates a diagonal covariance matrix V. Therefore, only the array of weights read in as part of the input and representing the diagonal elements of $W = V^{-1}$ are printed with their associated cases. If the error terms had not been assumed to be uncorrelated, the lower triangular position of W would also have been printed.

The minimum and maximum absolute residuals (min $|y_i - y_i|$ and max $|y_i - y_i|$) are also provided.

The user has the option of requesting confidence limits at the 100γ percent level, where $1-\gamma$ is specified by the user. These limits may be placed about the estimated values y_i for the original levels of X or for up to 100 other synthetic points.

At the same time, 100γ percent prediction limits, based on the predicted mean of m new observations at the same levels of X as used for the confidence limits may be requested. The value of m is also user provided. Pages five and six of the printout show 95-percent confidence and prediction limits using the original input values for the levels of X. The prediction limits are based on the predicted value of a single future observation at each level of X.

APPENDIX A

INPUT GUIDE FOR WEPOR AND WEPOR2

Input Guide for WEPOR and WEPOR2

Card No.	Variable	Description	Columns	Format
1	ITITLE	Title for run	1-80	8A10
2	NOBS	Number of observations NOBS > 0 data on cards NOBS < 0 data on TAPE8 (Must be attached prior to execution) WEPOR: $2 \leqslant \mid \text{NOBS} \mid \leqslant 750$ WEPOR2: $2 \leqslant \mid \text{NOBS} \mid \leqslant \mid 100$	1-5	I5
	KMAX	Desired degree of polynomial model	6-10	I5
	COPT	Confidence/prediction limit option COPT = 0 no intervals = 1 confidence intervals only = 2 confidence and prediction intervals Default: 0	11-15	I5
3 (used only if COPT = 1,2)	NPTS	Number of synthetic points for confidence/prediction limits NPTS = 0 use original x_i values NPTS $\leqslant$ 100	1-5	I5
	AR	AR = $(1 - \gamma)$ for 100γ percent limits $0 \leqslant$ AR $\leqslant 1.0$ Default: 0.05	6-10	F5.2

Card No.	Variable	Description	Columns	Format
	M	Number of future observations prediction limits based on Default: 1	10-15	I5
4	FORM1	Format used to read in (x,y) pairs	1-80	8A10
5	X	Independent variable level		FORM1
	Y	Dependent variable observation		FORM1
		(Repeat Card 5 as needed)		
6	FORM2	Format used to read in "weights"	1-80	8A10
7	WEPOR:W	Array of weights (diagonal elements of $W = V^{-1}$)		FORM2
	WEPOR2:V	Covariance matrix		FORM2
		(Repeat Card 7 as needed)		
8 (used only if COPT = 1,2 and NPTS $>$ 0)	XPTS	Synthetic points - - levels of independent variable		FORM1

APPENDIX B

SAMPLE INPUT AND OUTPUT FOR WEPOR AND WEPOR2

Sample Input Deck

DATA CARD LAYOUT
NUMBER ONE
NDW-NSWC/DL-10462/60 (REV. 04-75)

DATA CARD LAYOUT

```
PROJECTILE SEATING DISTANCE vs PERCENT GUN LIFE
   10    1    2
    0  .05
(2F5.2)
0.0  39.82
0.0  39.71
10.0 41.13
10.0 41.10
30.0 43.52
30.0 43.90
60.0 48.05
60.0 46.31
75.0 47.23
75.0 48.68
(10F7.2)
   12.5   12.5   10.0   10.0    5.0    5.0    1.4    1.4   1.25   1.25
```

S/N 0100-LL-L02-0436

GPO 915-355

```
PROGRAM WEPOR: WEIGHTED POLYNOMIAL REGRESSION
NAVAL SURFACE WEAPONS CENTER / DAHLGREN LABORATORY
APRIL,1981./

PROJECTILE SEATING DISTANCE VS PERCENT GUN LIFE


NUMBER OF OBSERVATIONS =    10
MAXIMUM DEGREE OF POLYNOMIAL =     1
CONFIDENCE INTERVAL OPTION (COPT) =     2


COPT = 0  (NO CONFIDENCE INTERVALS COMPUTED
       1  (CONFIDENCE INTERVALS PRINTED)
       2  (CONFIDENCE AND PREDICTION INTERVALS PRINTED)
M SET TO DEFAULT VALUE    1


          REGRESSION EQUATION FOR POLYNOMIAL OF DEGREE  1

               YHAT =      .39861E+02 +  (     .11767E+00)*(X**1)
```

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F TEST FOR LACK OF FIT
REGRESSION	104424.8918	2	52212.44591	
ERROR	6.740587209	8	.8425734011	
LACK OF FIT	2.866079708	3	.9553599028	1.23288
PURE ERROR	3.874507501	5	.7749015001	
TOTAL	104431.6324	10		

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARE	F STATISTIC
REGRESSION	104424.8918	2	52212.44591	
X**0	104106.3504	1	104106.3504	2880.45
X**1	318.5413838	1	318.5413838	378.058
ERROR	6.740587209	8	.8425734011	
TOTAL	104431.6324	10		

CASE NO.	X VALUE	WEIGHT	OBSERVED Y	PREDICTED Y	RESIDUAL
1	0.	12.50000	39.8200	39.8814	-.614065E-01
2	0.	12.50000	39.7100	39.8814	-.171406
3	10.0000	10.00000	41.1300	41.0581	.718818E-01
4	10.0000	10.00000	41.1000	41.0581	.418818E-01
5	30.0000	5.00000	43.5200	43.4115	.108458
6	30.0000	5.00000	43.9000	43.4115	.488458
7	60.0000	1.40000	48.0500	46.9417	1.10832
8	60.0000	1.40000	46.3100	46.9417	-.631677
9	75.0000	1.25000	47.2300	48.7067	-1.47674
10	75.0000	1.25000	48.6800	48.7067	-.267446E-01

MINIMUM ABSOLUTE RESIDUAL = .267446E-01
MAXIMUM ABSOLUTE RESIDUAL = 1.47674

95% CONFIDENCE INTERVALS FOR SYNTHETIC POINTS

CASE NO.	X VALUE	LOWER CONFIDENCE BOUND	PREDICTED Y	UPPER CONFIDENCE BOUND
1	0.	39.545	39.881	40.218
2	0.	39.545	39.881	40.218
3	10.000	40.779	41.058	41.337
4	10.000	40.779	41.058	41.337
5	30.000	43.061	43.412	43.762
6	30.000	43.061	43.412	43.762
7	60.000	46.247	46.942	47.637
8	60.000	46.247	46.942	47.637
9	75.000	47.815	48.707	49.598
10	75.000	47.815	48.707	49.598

95% PREDICTION LIMITS FOR SYNTHETIC POINTS
BASED ON THE PREDICTED MEAN VALUE OF 1 FUTURE OBSERVATIONS

CASE NO.	X VALUE	LOWER PREDICTION LIMIT	PREDICTED Y	UPPER PREDICTION LIMIT
1	0.	37.738	39.881	42.025
2	0.	37.738	39.881	42.025
3	10.000	38.923	41.058	43.193
4	10.000	38.923	41.058	43.193
5	30.000	41.266	43.412	45.557
6	30.000	41.266	43.412	45.557
7	60.000	44.714	46.942	49.170
8	60.000	44.714	46.942	49.170
9	75.000	46.410	48.707	51.003
10	75.000	46.410	48.707	51.003

APPENDIX C

SAMPLE PLOTS PRODUCED BY PROGRAM LIMITS

Program LIMITS uses the graphics package DISSPLA (Reference 3 in text) to plot the confidence and prediction limits generated by programs WEPOR and WEPOR2. Local files produced by these two programs and used as input for LIMITS are TAPE10 (raw data), TAPE11 (confidence limits), and TAPE12 (prediction limits). Figures C-1 and C-2 were drawn using the results of the example discussed on page 9.

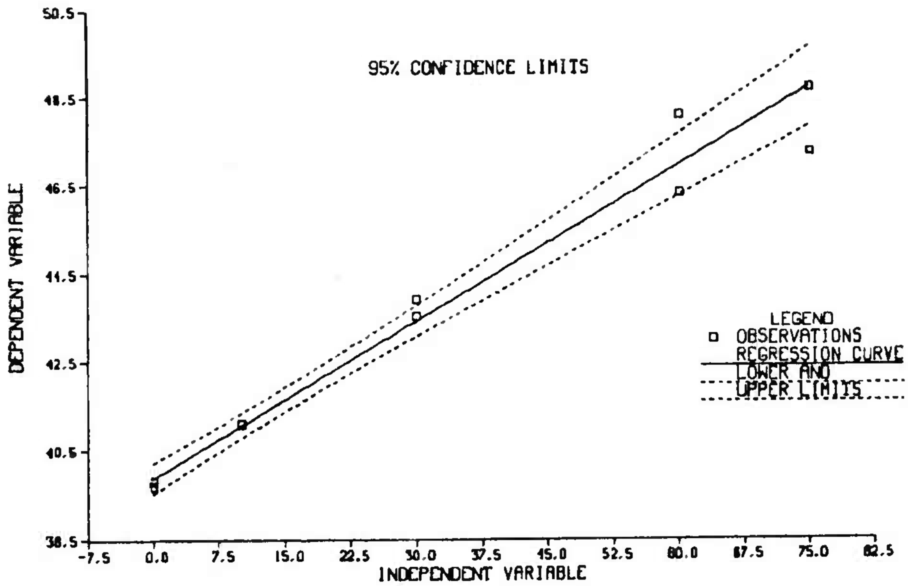

Figure C-1. 95-Percent Confidence Limits: Projectile Seating Distance Expressed as a Function of Percent Gun Barrel Life Expended

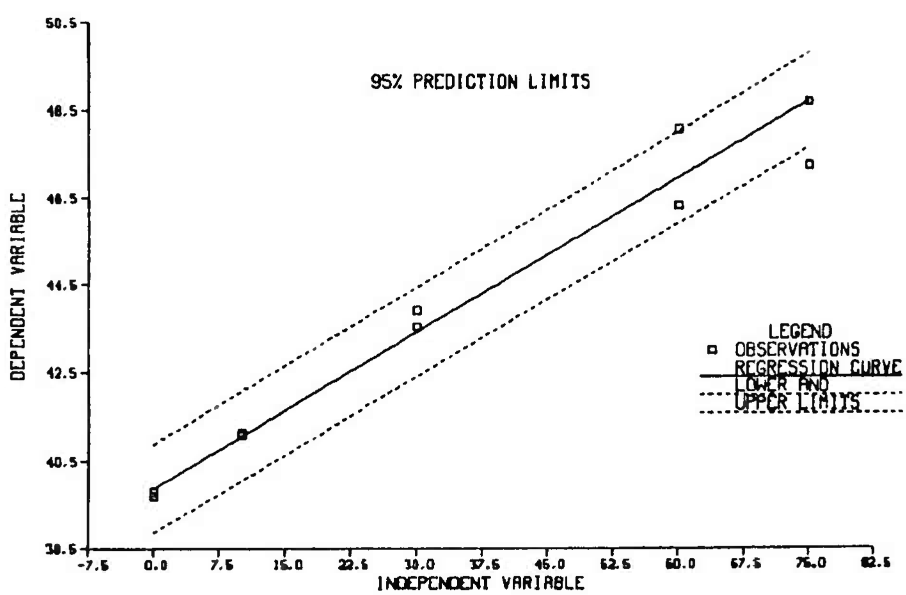

Figure C-20. 95-Percent,Prediction Limits: Projectile Seating Distance Expressed as a Function of Percent Gun Barrel Life Expended

DISTRIBUTION

Commander
Naval Air Development Center
ATTN: Technical Library
Warminster, PA 18974

Commander
U. S. Army Missile Research and
Dvelopment Command
ATTN: Technical Library
Redstone Arsenal, AL 35809

Director
Naval Research Laboratory
ATTN: Code 5314 (Mr. Duncan)
Technical Information Division
Washington, DC 20375

Office of Naval Research
Department of the Navy
ATTN: ONR-436
Washington, DC 20360

Commanding Officer
U. S. Air Force Armament Laboratory
ATTN: Technical Library
Eglin Air Force Base, FL 32542

Commanding Officer
Naval Weapons Center
ATTN: Technical Library
Code 324C (Kurotori)
China Lake, CA 93555

Los Alamos Scientific Laboratory
P. O. Box 1663
ATTN: Library
Los Alamos, NM 87544

DISTRIBUTION (Continued)

Commander
U. S. Army Armament Research and
Development Command
ATTN: DRDAR-LCS-E (Dr. Einbinder)
Technical Library
Dover, NJ 07801

Commander
U. S. Army Armament Material
Readiness Command
ATTN: DRSAR-PES (Mr. Michels)
Technical Library
Rock Island, IL 61299

Director
U. S. Army Ballistic Research Laboratory
ARRADCOM
ATTN: DRDAR-BLY (Mr. Danish)
Technical Library
Aberdeen Proving Ground, MD 21005

Oklahoma State University
Field Office
P. O. Box 1925
ATTN: Mrs. Peebles
Eglin Air Force Base, FL 32542

Director
U. S. Army Material System Analysis Activity
ATTN: DRXSY-GS (Mrs. Ritondo)
Technical Library
Aberdeen Proving Ground, MD 21005

Director
Naval Ship Research and Development Center
ATTN: Technical Library
Washington, DC 20034

U. S. Army Mathematics Research Center
University of Wisconsin
ATTN: Technical Library
Madison, WI 54306

DISTRIBUTION (Continued)

Superintendent
U.S. Naval Postgraduate School
ATTN: Technical Library
Monterey, CA 93940

Superintendent
U. S. Naval Academy
ATTN: Library
Annapolis, MD 21402

Commanding Officer
U. S. Army Research Office
Durham, NC 27706

Department of Statistics
VPI&SU
ATTN: Statistical Laboratory
Blacksburg, VA 24060

Department of Statistics
North Carolina State University
Raleigh, NC 27105

Department of Statistics
University of Iowa State
Iowa City, IA 50703

Virginia Commonwealth University
Biometry Division
ATTN: Dr. Hans Carter
Richmond, VA 23219

Department of Statistics
Purdue University
ATTN: Dr. George McCabe, Jr.
Lafayette, IN 47905

DISTRIBUTION (Continued)

Defense Technical Information Center Cameron Station Alexandria, VA 22314	(2)
Library of Congress ATTN: Gift and Exchange Division Washington, DC 20540	(4)

Local:

E31 (GIDEP)	
E411 (Hall)	
E431	(10)
K	
K10	
K106	(25)
K11	
K11 (Phillips)	
K33 (Morris)	
K34	
K40	
K41	
K42	
K43	
K44	(4)
K50	
K52 (Farr)	(4)
G	
G11	
G12	
G12 (Hinkle)	
G13	
G22 (Clawson)	
G25 (Jennings)	
G32 (Seidl)	
G32 (Monn)	
G60 (Jablovski)	
G61	
R11	
R13	(2)

DISTRIBUTION (Continued)

R14	(3)
R16	(4)
R32	(5)
R34	(6)
R44	(7)